

AZ ALAKTANIG ÉS TOVÁBB: KORCHMÁROS, KOCSMÁROS, KORCHOMÁROS ÉS TÁRSAI

MORFOLÓGIAILAG ELEMZETT TÖRTÉNETI MAGÁNÉLETI ADATBÁZIS

DÖMÖTÖR ADRIENNE

Marótiné Korchmáros Valériának korábban tanítványa, később kollégája lehettem Szegeden – mindkét helyzetben igen sokat tanultam tőle. Kritikája, tanácsai a mai napig fontosak számomra. Volt alkalmam megtapasztalni azt is, hogy az „és mit csinálsz mostanában?” típusú kérdések nála nemcsak a kötelező udvariasság jelei – érdeklődésére valódi, szakmai választ vár. Köszöntésekor – amely meglepetésnek volt szánva – szeretnék elébe menni ennek a kérdésnek. S mivel a tanszék meghívása morfológiai előadásra szólt, a két dolgot össze is kapcsolnám.

Nem egészen három hónappal ezelőtt, 2010 szeptemberében indult egy négy éves munkálat, amelynek középpontjában a morfológia áll – ahogyan erről előadásom alcíme is árulkodik. A *Morfológiailag elemzett nyelvtörténeti korpusz a magánéleti nyelvhasználat köréből*¹ című projektum tervét, céljait és munkafázisait szeretném röviden bemutatni.

Elektronikus adatolás

A nyelvtörténet – természetéből következően – adatigényes tudományág: korábbi korok nyelvi változásait csak megfelelő mennyiségű és gondosan kiválasztott forrásokból összegyűjtött anyag birtokában lehet felrajzolni (már amennyiben nyelvmélekkel rendelkező időszakról van szó). Elektronikus korpuszok segítségével a felhasználók összehasonlíthatatlanul könnyebben és gyorsabban juthatnak adatokhoz, mint korábban, amikor a kutatásokat – meglehetősen idő- és energiaigényes módon – előzetes anyagválogatásnak és -vizsgálatnak, valamint saját adatgyűjtésnek (cédulázásnak) kellett megelőznie.

Ez a könnyebbség minden eddiginél nagyobb vonzerőt kölcsönöz az adatoknak, ami talán hozzájárulhatott ahhoz is, hogy a – korábban kizárólag az introspekció létjogosultságát elfogadó – generatív szemléletű kutatások is egyre inkább érdeklődést mutatnak a gyűjtött adatok iránt. Az adatbázisok létrehozása visszamenőleg is jelentősen megnöveli a korábban elvégzett filológiai munka értékét, hiszen korpuszt építeni csak a már rendelkezésre álló, szakszerű kritikai szövegkiadásokból lehet, illetve érdemes. A digitális változatok ugyanakkor nem teszik feleslegessé a papíralapú publikációkat sem, hiszen átfogó anyagismeret és a szövegkiadások filológiai apparátusának felhasználása nélkül, pusztán kiragadott korpuszadatokból aligha születhetnek megbízható elemzések.

¹ OTKA 81189. sz. projektum. Vezeti: Dömötör Adrienne.

A munkálat kezdete; az adatbázis alapanyaga

A Nyelvtudományi Intézet korábbi Nyelvtörténeti Osztályából létrejött háromtagú kutatócsoport kezdetben azt tervezte, hogy ómagyar kori korpuszt épít a kódexek és a kisebb nyelvemlékek anyagából; egyrészt a TNyt. és a kódexkiadások munkálatai során felgyülemlett tapasztalatokra, másrészt pedig a kiadott kódexeknek a gépeinken lévő digitális szövegváltozataira alapozva. A csoport (további intézeti munkatársak bevonásával) próbajelleggel hozzá is kezdett a munkához; először a kézi annotálás feladataival és lehetőségeivel próbálkozva, majd a Novák Attila számítógépes nyelvész által javasolt „szövegnormalizálás” módszerével kísérletezve. (Ez utóbbiról később még bőven lesz szó.) Néhány szövegrészlet elkészülte után azonban, amikor tudomásunkra jutott, hogy az intézet egyik kutatócsoportja megnyert egy generatív nyelvtörténeti pályázatot, amelynek az ómagyar kori korpusz megépítése is a részét képezi, másik alapanyagot választottunk (meglévő digitális szövegváltozatainkat pedig átadtuk annak a projektumnak).

Az új választás során kijelölt forráscsoport jelentőségét az adja, hogy a magánéleti nyelvhasználatba – és a mögötte kitapintható nyelvi tendenciákba, szabályokba – nyújt betekintést. A magánéleti regiszterhez legközelebb álló műfajokként a levelek és a periratok (tanúvallomások) kínálkoznak arra, hogy korábbi századok mindennapi nyelvhasználatát tanulmányozhassuk. Ezért az ó- és középmagyar kori levélanyag, valamint az 1772-t megelőzően keletkezett perszövegek képezik az épülő adatbázis anyagát, ami az ómagyar korból csekély, ám a középmagyarból annál tetemesebb anyagmennyiséget jelent. A pályázati időszak alatt természetesen nem nyílik mód a teljes elérhető anyag feldolgozására; az adatbázis azonban bármikor tovább bővíthető.

Az anyag kijelölése után először megkíséreltünk némi gyakorlatot szerezni a korpuszepítés újfajta módszerében, hiszen a Novák Attila által ajánlott, majd közösen, sok-sok aprómunkával kidolgozott eljárásnak ez volt az első próbája. (Korábbi próbamunkánkból két részlet – az egyik még a kódexekből, a másik már a középmagyar magánéleti szövegekből – a www.nytud.hu-n, a nyelvtörténeti munkacsoport honlapján látható.) Ezt követte a pályázat beadása, majd a kedvező döntés után – immár külső munkatársak, elsősorban PhD-hallgatók bevonásával – a négy millió karakter nagyságúra tervezett korpusz megépítésének kezdete.

Digitális szöveggyűjtemény

Az adatbázis elsősorban a történeti morfológiai, lexikológiai, szociolingvisztikai kutatásokhoz kínál anyagot, de segítségül szolgál a történeti mondattani és pragmatikai vizsgálatokhoz is, ahogyan mindezen területek oktatásához is használható. Az anyag összeállítása során jelentős figyelem jut a szociolingvisztikai szempontoknak: a perek szövegei változatos időkort és területi megoszlást képviselnek, a leveleknél pedig további szempontok is figyelmet kapnak: az író társadalmi státusza, neme, a címzetthez fűződő kapcsolata és a levél létrejöttének módja (saját kezű-e vagy sem).

A nyelvtörténeti adatok publikálása terén már az is jelentős előrelépés volna, ha pusztán elektronikusan közzétennénk a szövegeket. Meghatározott betűsorokat így is nagy mennyiségben lehetne kigyűjteni az anyagból. Teljes listákat is össze tudna állíta-

ni az, aki járatos a keresőprogramok kínálta lehetőségekben, és otthonosan mozog a tudományterületen is, vagyis fel tudja mérni, milyen alakokra kell rákeresni, amikor egy adott szó összesféle előfordulását meg akarja kapni. Ehhez a vizsgálni kívánt jelenség előzetes tanulmányozására éppúgy szükség van, mint az adott szöveg hangjelölési, helyesírási sajátosságainak felmérésére.

Ha például a *kocsmáros* szó boszorkányperek-beli előfordulásait szeretné valaki összegyűjteni, az anyag tanulmányozása után legalábbis a következő formákra kell számítania (és mellettük természetesen előfordulhatnak előre kalkulálhatatlan, egyéni írásváltozatok, íráshibák is):

<i>korchmáros</i>	<i>korchomáros</i>	<i>kochmáros</i>	<i>kochomáros</i>
<i>korchmaros</i>	<i>korchomaros</i>	<i>kochmaros</i>	<i>kochomaros</i>
<i>korcsmáros</i>	<i>korcsomáros</i>	<i>kocsmáros</i>	<i>kocsomáros</i>
<i>korcsmaros</i>	<i>korcsomaros</i>	<i>kochmaros</i>	<i>kocsomaros</i>
<i>korczmáros</i>	<i>korczomáros</i>	<i>koczmaros</i>	<i>koczomáros</i>
<i>korczmaros</i>	<i>korczomaros</i>	<i>koczmaros</i>	<i>koczomaros</i>
<i>korcmáros</i>	<i>korcomaros</i>	<i>kocmáros</i>	<i>kocomáros</i>
<i>korcmaros</i>	<i>korcomaros</i>	<i>kocmaros</i>	<i>kocomaros</i>

Egy elemzetlen szövegtörzshoz tehát elsősorban a magyar nyelvtörténet szakemberei tudnának megbízható gyűjtéseket végezni. S nekik is további – esetleg jelentős – kiegészítő munkát jelentene a homonim betűsorok kiszűrése a találati listából (gondolhatunk itt nemcsak a *vár* ige és *vár* főnév típusú homonímiákra, hanem a grammatikailag létrejött homonim betűsorokra is, mint például a *nevet* ige és *nevet* főnév + tárgyrag esete).

Morfológiailag elemzett adatsor

A munkát az elmondottak miatt lényegesen többre törekszik annál, minthogy pusztán összeállítson egy elektronikus szöveggyűjteményt. A morfológiai feldolgozásnak a felhasználó szempontjából az a felbecsülhetetlen jelentősége, hogy célirányos keresésre ad módot. Így már nemcsak adott betűsorokat lehet lekérdezni, hanem lehetőség van morfológiai kategóriákra is rákeresni. Abban az esetben pedig, ha a felhasználó betűsorokra keres rá, elegendő megadnia az elem mai formáját ahhoz, hogy az összes korabeli előfordulást megkapja. Ezt az teszi lehetővé, hogy az eredeti elemen kívül a mai sztenderd változatuk is be van kódolva. A rendszer minden esetben a megcélzott elem(csoport)ot listázza ki, homonimák nélkül. A gyűjteményt így a kifejezetten magyar nyelvtörténeti képzettségű felhasználókon kívül mások is – egyéb nyelvészeti részdiszciplínák kutatói, egyetemi hallgatók, külföldi nyelvészek stb. – megbízhatóan kezelhetik adatforrásként.

Nézzünk először egy betűsorra való rákeresést, illetve rögtön ennek az eredményét! Ha például a *kocsmáros* szó előfordulásait szeretnénk összegyűjteni, elegendő megadnunk ezt az egy alakot ahhoz, hogy az eredetiben az alábbi adatokat is megkapjuk:

*e fatensnek anya Baksa Eva **korcsmáros**néval öszvő szolakodván,
azon szollakodásajért Baksa Eva meg fenyegette e fatens Annyát*

*azon Talban valami bogarak es soh leven
az Varas **korczmaros** haza előtt hintete az Pincze elől*

*Nagi Hoflany **korcomaros** Faingli Martony huty utan vallia,
hogy az eő felesége Matthias Swadlsek kurvaja let volna*

Lássunk példát morfológiai kategóriákra való rákeresésre is – ezúttal már a lekérdezés folyamatával együtt! Ha például azokat a szerkezeteket szeretnénk megkapni, amelyekben egy főnévnek a létige befejezett melléknévi igeneve és egy melléknév is a jelzője, a keresőbe a *V PartPrf Adj N* kombinációt kell beírunk. (A nemzetközi felhasználhatóság érdekében a morfológiai kategóriák rövidítései a nemzetközi sztenderdet követik; feloldásukat egy jegyzék tartalmazza. A lentebbi ábra keresőkérdésében szereplő *L a lemmá*-t, a *C a category tag*-ot [=címke] jelöli.) A „Mehet” gombra kattintva környezetükkel együtt kilistázódnak a kívánt felépítésű szerkezetek, így például a következő adatsor is:

TMK Történeti Magánéleti Korpusz lekérdezőfelület

Lekérdezés: [L~van@C~V~PartPrf Adj(0)-] N]

Megjegyzés: []

Adatbázis: [perlev] Szöveggjellemzők: []

Mehet v1.0.6 - 2012.09.11. - [] [] []

való	bosszúját	mindjárt	másik	éjszakán	álmi	kíváncsn	elannyira	megnyomta,	hogy	<ha	már	a	házi	cselédék	fel	ne	ketek	volna,>	
Adj	N.PxS3	Adv	AdjPro	N.Sup	V.Inf	V.PartAdv=Ván	AdvPro	V.Ftx.V.Past.S3	Def	C	C	Adv	Det	Adj	N.PI	V.Ftx	Adv	V.Past.P3	V.Cond

el	kölletett	volna	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,
el	kelletett	volna	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,
el	kelletett	volna	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,	fulladni,
VPfx	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3	V.Cond	V.Past.Past.S3

[3] Bosz. 464. Vao megye, 1648. > > (NE) - 254381

hatAd-7mom	Ezen	fatens	hallotta	ennyihán	izben	is	a	maga	Feleséghit,	hogy	az	előbeni	Urának	össze	lívén	halála	előt	circiter	két	hettel	kucsolodva
hatAd-7mom	Ezen	fatens	hallotta	ennyihán	izben	is	a	maga	Feleséghit,	hogy	az	előbeni	Urának	össze	lívén	halála	előt	circiter	két	hettel	kucsolodva
Det	Det	N	V.Past.S3	Adv	Adv	Adv	Adv	Adv	N.PxS3	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	

a	szája,	mindenkor	nem	egyébre,	hanem	Szabó	Mihály	gyanakodott,	a mint	hogy	ugyan	akkor	emített	Urának	(aki	is	a	száját	fel	nem
a	szája,	mindenkor	nem	egyébre,	hanem	Szabó	Mihály	gyanakodott,	a mint	hogy	ugyan	akkor	emített	Urának	(aki	is	a	száját	fel	nem
a	száj	mindenkor	nem	egyébre,	hanem	Szabó	Mihály	gyanakodott,	a mint	hogy	ugyan	akkor	emített	Urának	(aki	is	a	száját	fel	nem
Det	N.PxS3	AdvPro	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	

nithatván	nyithatván	nyit	csak	a	foghai	álmal	kintelenített	szolani)	Kapuvárt	akkor	foghsagban	lívő	Szighety	névő	völt	Babothi	Korcsmáros	tanácsolta	azt,
nyit	nyit	nyit	csak	a	fog	álmal	kintelenített	szolani)	Kapuvárt	akkor	foghsagban	lívő	Szighety	névő	völt	Babothi	Korcsmáros	tanácsolta	azt,
V.Mod	V.Mod	V.Mod	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv	Adv

hogy	rostályt	a tűze	tegye,	és	azon	két	mosófát	pitson	meg,	az	után	meg	jóttul,							
hogy	rostályt	a tűze	tegye,	és	azon	két	mosófát	pitson	meg,	az	után	meg	jóttul,							
C	Det	N.Acc	Det	N.Sub	V.Subj	S3	Def	C	Det	Pro	Q	N.Acc	V.Subj	S3	Def	C	Det	Pro	Q	N.Acc

1. ábra: A lekérdezőfelület működése

Az elemzőprogram tehát megadja és kereshetővé teszi a szótöveket, azok szófa-ját és a szótőn fellelhető grammatikai elemeket is – azaz hogy a szónak milyen számú,

személyű, idejű, milyen névszóraggal ellátott stb. – alakjával van dolgunk (a képzőktől megfosztott abszolút szótövekig azonban nem megy vissza).

Ezt az automatizált munkaszakaszt – programozói oldalról éppúgy, ahogyan a korpuszt építő nyelvtörténész oldaláról – számos, aprólékosan elvégzett munkafolyamatnak kell megelőznie és követnie. A továbbiakban ezekről a munkafázisokról lesz röviden szó, miközben kialakul a válasz az előző ábra láttán bizonyára felmerülő kérdésre is, hogy milyen sorok és milyen okból állnak itt egymás alatt.

A korpuszpépítés folyamata

Az első lépés: a szövegeket elektronikusan olvashatóvá kell tenni. (Ennek részletezésétől most tekintsünk el a kíváncsú rövidség érdekében!) A munka lényegi szakasza, a feldolgozhatóvá tétel ezután kezdődik. Annak érdekében, hogy a mai magyar szövegek feldolgozására kifejlesztett elemzőprogram nyelvtörténeti szövegeken is működtethető legyen, először egy új szövegváltozatot kell létrehozni.

Ez a munkaszakasz az úgynevezett normalizálás, amely a szöveget a mai sztenderd nyelvváltozathoz közelíti, és alkalmazkodik a mai hangjelölés és helyesírás szabályaihoz – így válva alkalmassá arra, hogy közvetítő szerepet töltsön be a történeti szöveg és az elemzőprogram lehetőségei között. Megfelelő szabályok szerint át kell tehát alakítani az eredeti szövegek összes mondatát, majd begépelni, és – több körben – ellenőrizni, javítani, a vitás kérdéseket ezenközben tisztázni.

Ahhoz, hogy a morfológia szintjén ne legyen eltérés az eredeti és a normalizált szövegváltozat között, a legfontosabb szabály az, amelyet házi használatra a „morféma-megmaradás törvényé”-nek neveztünk el. A mai megfelelővel nem rendelkező – vagy eltérő használatú – morfémáknak meg kell maradniuk eredetiben a sztenderdizált szóalakon is; ellenkező esetben elvesznének az elemzésből. Például a *nekije*-ből nem lehet *neki*, a *dolgoztanak*-ból *dolgoztak*, a *Pécsé mene*-ből *Pécsre ment*. Ezekben az esetekben az elemzőprogrammal kapcsolatban adódik feladat: „meg kell tanítani” a számára új elemek, szokatlan elemkombinációk felismerésére. A megtanítandó tövek között például olyanok fordulnak elő, mint *kapufőle*, *lemonya*, *pázsitdisznó*, *tyúkmonysütte*, *csentereg*. Számos esetben történeti, etimológiai és tájszótárak, valamint történeti grammatikák tanulmányozása előzi meg a döntést, hogy például helyes megoldás-e, ha a *moh*-ból *moha*, a *másüvé*-ből *máhová*, a *májog*-ból *nyávog* lesz.

Ebben a munkaszakaszban – tagmondatokra tagolva, és az eredeti alá a normalizált változatot beírva – az alábbihoz hasonlóan festenek a szövegek (némi leegyszerűsítve, hiszen a helyes kódolás érdekében az eredeti változatban is számos technikai jellegű jelölést kell alkalmazni):

Boszorkányságához egyebet nem tud
Boszorkányságához egyebet nem tud,
hanem Tavalyi esztendőben itt Balkányban lévén Valyi János uramnak egy
korcsmarossa,
hanem tavalyi esztendőben itt Balkányban lévén Valyi János uramnak egy
kocsmárosa,

*és annak lévén egy szolgálcskája mint egy tizen öt esztendőś,
és annak lévén egy szolgálcskája, mintegy tizenöt esztendőś,
és ezen Fatens oda menvén
és ezen fatens odamenvén,
már akkor az szolgál igen igen rosszul vólt
már akkor a szolgál igen-igen rosszul vólt.*

A program a sorpárok alá beilleszti a morfológiai elemzéseket (együttal átalakítja az eredeti szövegbe beszúrt technikai jelöléseket is). A háromféle adatsor – az eredeti szöveg, a normalizált változat és a szavakhoz rendelt morfológiai elemzés – együttesen is megjeleníthető; ahogyan kereséskor is lekérdezhető a három sor bármelyike külön-külön vagy együtt is. Az alábbi ábrán együtt látható a három adatsor. (A látszólagos különbség az előző ábrától abból adódik, hogy ott a morfológizált sor két alsorra törölve volt látható.)

struc:addr										
Az	en	İerelmes	Atiamfjanak	Wardaj	Kat	aİonnak	adassık.			
Az	én	szerelmes	atyámfjának,	Várdaj	Kata	asszonynak	adassék.			
az[Det]	én[N Pro.S1]	szerelmes[Adj]	atyja+fia[N.PxS1.Dat]	Várdaj[N]	Kata[N]	asszony[N.Dat]	ad[V.Pass.Subj.S3]			
struc:addr										
Kóİőnetem		magam	ajanlása	után,						
Kőszőnetem,		magam	ajánlása	után,						
kőszőnet[N.PxS1]		maga[N Pro.S1]	ajánlás[N.PxS3]	után[PP]						
Isten	agion	mjnden	jokat	jó	egjsiget	Henele-mjnd	az	apro	gjermekekkel	egjetemben.
Isten	adion	minden	jókat,	jó	egészséget	mind	az	apró	gyermekkel	egyetemben.
isten[N]	ad[V.Subj.S3]	minden[Det]	jó[Adj.Pl.Acc]	jó[Adj]	egészség[N.Acc]	mind[Adv]	az[Det]	apró[Adj]	gyermek[N.Pl.Ins]	egyetem[N.Ine]
Istennek	hala									
Istennek	hála,									
isten[N.Dat]	hála[N]									
mj	egjsigben	uagıunk.								
mi	egészségben	vagyunk.								
mi[N Pro.P1]	egészség[N.Ine]	van[V.P1]								
Mi nekünk	mjnd	ez	jdeıg	ualo	jargalasunk	sem	ır	egj	barackot,	
Minekünk	mind	ez	ideıg	való	járkálásunk	sem	ér	egy	barackot,	
mi[N Pro.Dat.P1]	mind[Adv]	ez[Det Pro]	ideı[N.Ter]	való[Adj]	járkálás[N.PxP1]	sem[Adv]	ér[V.S3]	egy[Det]	barack[N.Acc]	
chak	ıjaban	hurcolnak	faraİtanak	bennünket	az	sok	saron,	hegjen,	ıizen.	
csak	ııába'n	hurcolnak,	fárasztanak	bennünket	a	sok	sáron,	hegyen,	vızen.	
csak[Adv]	ııába[Adv]	hurcol[V.P3]	fáraszt[V.P3]	mi[N Pro.P1.Acc]	a[Det]	sok[Q]	sáır[N.Sup]	hegy[N.Sup]	vız[N.Sup]	

2. ábra: Az adatsorok együttes megjelenítése

Ezen együttállítás létrehozatala a korpuszépítő számára további – itt nem részletezendő – nagy mennyiségű, változatos feladatot jelöl ki. A napi szinten megválaszolható szakmai kérdések sorában a legelőkelőbb helyet a normalizálás menete közben felvetődő lehetőségek mérlegelése foglalja el. Igyekszünk a lehető legnagyobb következetességgel eljárni; ennek érdekében a normalizálási szabályzat folyamatosan alakul, bővül.

De hasonlóképpen sok rákérdezést igényel az elemzőprogram által elvégzett munka is. Egyrészt a megoldásait ellenőrizve időről időre pontosítani, javítani, tanítani

kell a programot. Tanácsos ehhez számos nyelvi kérdést újragondolni, hiszen ebben a munkában nem lehet átmeneti kategóriákat, határeseteket megadni.

És végig kell vinni egy mechanikusabb munkafolyamatot is. Homonim alakok esetében az elemző az összes elvi lehetőséget felajánlja, hiszen – az emberi elmétől eltérően – önállóan nem tud szelektálni a homonimák között. Ez a jelenség nagyon nagy számban érinti az alakokat, és a lehetőségek sora sokszor igen tekintélyes. Például a *megérttettem* szóalak elemzéseként tíz változat jelenik meg:

megért [V.Past.S1]
megértet [V.Past.S1]
megértetik [V.Past.S1]
megért [V.Past.S1.Def]
megért [V.Past.S1.Def?]
megértet [V.Past.S1.Def]
megértet [V.Past.S1.Def?]
megértetik [V.Past.S1.Def]
megértetik [V.Past.S1.Def?]
megért [V.PartPrf.PxS1]

Ezekon a helyeken a munkatársaknak kézi eljárással kell az egyértelműsítést elvégezni, azaz egy szövegdoboz megnyitásával a felkínáltak közül az adott kontextusban érvényes megoldást kiválasztani; a *kegyelmed leveléből nyilván megérttettem, hogy...*-ot tartalmazó szöveghelyen például a *megért [V.Past.S1]*-et megjelölni. Az alábbi ábrán egy jóval egyszerűbb választás szerepel példaként (a *lopó* főnév + többesszám-jel kódot kell választani a melléknévi igenév, illetve a többes szám első személyű igealak ellenében).

valakinek Valakinek vala+ki[N Pro.Dat]	Gyermekeket gyermekét gyermek[N.PxS3.Acc]	vestete vesztette veszt[V.Past.S3.Def]	avagi avagy avagy[C]	Marhat marhát, marha[N.Acc]
avagi avagy avagy[C]	Mezőben mezőben mező[N.Ine]	Gonoszt gonoszt gonosz[N.Acc]	czelekedé; cselekedé? cselekszik[V.Ipf.S3]	
avagi Avagy avagy[C]	Mások mások, más[Adj/N Pro.Pl]			
az kik akik a+ki[N Pro.Rel.Pl]	loptanak loptanak lop[V.Past.P3=tAnAk]	es és és[C]	Tolvajloptanak tolvajloptanak, tolvajlik[V.Past.P3=tAnAk]	
es és és[C]	or Gazdak orgazdák orgazda[N.Pl]	volnanak volnának, van[V.Cond.P3]		
avagi avagy avagy[C]	lophok lopók lopó[N.Pl]	volnanak. volnának? van[V.Cond.P3]		
!let14.-test !let14.-test	lopó[N.Pl] lop[V.PartPrs.Pl] lop[V.Ipf.P1.Def]			

3. ábra: Példa a kézi egyértelműsítésre

Felhasználás

Az adatbázis elkészítése annál több energiabefektetést igényel, minél megbízhatóbban szeretné kiszolgálni majdani felhasználóit. Az előzőekben talán sikerült érzékelteni, hányféle munkafolyamat és milyen mennyiségű munka rejlik amögött, hogy a felhasználó pár kattintással akár jelentékeny mennyiségű adathoz juthasson.

Az elkészült adatbázis bárki számára – szakembereknek és érdeklődőknek – szabadon hozzáférhető lesz. Használatát egy bevezető útmutató segíti majd, amelyhez (ahogy erről már szó volt) a grammatikai kategóriák nemzetközi rövidítéseinek feloldása csatlakozik.

A projektum eredményeként létrejön az első magyar nyelvű történeti magánéleti adatbázis. A korpuszt sokoldalúan lehet alkalmazni: az érdeklődők használhatják pusztán szövegolvasásra és -értelmezésre, az egyetemi oktatók és hallgatók számára munkanyagként szolgálhat, a kutatóknak pedig – a munkálat fő célja szerint – adatforrást kínál, mégpedig a nyelvnek egy olyan rétegéből, amelynek igen nagy a jelentősége a nyelvtörténeti vizsgálatok számára.

MOPHOLOGY AND BEYOND: *KORCHMÁROS, KOCSMÁROS, KORCHOMÁROS* AND THEIR VARIANTS
MORPHOLOGICALLY ANALYSED DATABASE OF HISTORICAL TEXTS
REPRESENTING INFORMAL LANGUAGE USE

ADRIENNE DÖMÖTÖR

My paper describes a project which is in progress in the Research Institute for Linguistics of the Hungarian Academy of Sciences. The project aims at building a morphologically annotated historical corpus that intends to represent the daily vernacular of speakers of Old and Middle Hungarian, covering a period of three centuries from the end of the 15th century to the end of the 18th century. The corpus will comprise samples of private correspondence and depositions of witnesses in trials. The present paper introduces the method of the work and describes the different steps of the building process.